Combination of Dimensionality Reduction Techniques for Face Image Retrieval: A review

Fousiya K.K.
M.Tech Scholar, MES College of Engineering, Kuttippuram, Kerala

Jahfar Ali P.
Asst. Professor, MES College of Engineering, Kuttippuram, Kerala

ABSTRACT
Face image retrieval systems have attained much importance in recent times, due to many real time application demands it. Normally, image retrieval has become a challenging issue in the real world applications because the image collections are often high dimensional. This higher dimensionality results in a degradation of the classification accuracy of the classifier. Face recognition is the most important process of the face retrieval systems. The major step of face recognition is feature extraction and classification. The performance of a face recognition system highly depends on the way in which the features are extracted and the classification of them to the appropriate group. By providing better dimensionality reduction and higher class discrimination prior to the classification process, a classifier must provide higher classification accuracy. So, dimensionality reduction has a significant role in the classification process. This preprocessing step must provide a better dimensionality reduction of high dimensional data as well as high discrimination between classes. In this paper an attempt has been made to summarize some of the existing dimensionality reduction methods and its relevance to improve the classification accuracy of the existing face retrieval systems. The proposed method has improved the classification accuracy into a higher rate.

General Terms
Pattern Recognition; Face recognition

Keywords
Face image retrieval; Dimensionality reduction; PCA; LDA; SVM.

1. INTRODUCTION
A face image retrieval system retrieves the relevant face images from a large face database, which are similar to a query face image. Face recognition system is a major issue in applications such as security systems, visual surveillance and credit card verification. Face recognition is the process of recognizing human faces with many variations in facial appearances such as facial expression, poses and facial details. Studies show that, from a machine learning perspective image retrieval can be viewed as a problem of classifying images into the appropriate classes (Serigo A.A, 2002). Therefore, face image retrieval directly depends on the classification of the face images to the proper classes and so, face recognition is the major step in the face retrieval system. So, the retrieval rate directly depends on the classification accuracy.

Normally, the images including digital photographs, scientific images, medical images, hyper spectral images etc. are high dimensional and there exist hundreds or even thousands of features to represent an image. When use with machine learning and data mining algorithms, these larger number of features lead to a problem of higher dimensionality in image retrieval process. This curse of dimensionality leads to the degradation of performance as well as efficiency of the system. For the efficient processing of high dimensional images, its dimensionality has to be reduced without a loss in the original properties of the image represented in the high dimensional space.

Face images are often high dimensional. So, face recognition has to address the dimensionality reduction problem because, a face image having m*n pixels is represented by a vector in R^{m*n} space for computational purposes. This m*n dimensional space is too large for recognition process. When the number of images in the data set increases, the complexity of representing the data set also increases. Thus, for solving the curse of dimensionality, various dimensionality reduction methods have been applied as a preprocessing step.

1.1 Dimensionality Reduction
Dimensionality reduction is the transformation of high dimensional patterns into a lower dimensional representation. The goal of the dimensionality reduction is to preserve the local structure of the original high dimensional space. Dimensionality reduction removes the irrelevant and redundant features, and extracts a small number of relevant features. The minimum number of parameters which are required for representing the properties of the data is known as the intrinsic dimensionality. So, the dimensionality of the reduced representation should correspond to the intrinsic dimensionality. The lower dimensional representation provides the efficient visualization of the data. Similarly, the analysis of the high dimensional data can be done efficiently by projecting them into a lower dimensional representation. For example, the images produced by the hyper spectral satellites contain hundreds of spectral bands. For remote sensing applications, these higher bands are a major problem and the preprocessing requires a reduction in the number of bands, which is shown in Figure 1 (G. Chen et. al, 2007)

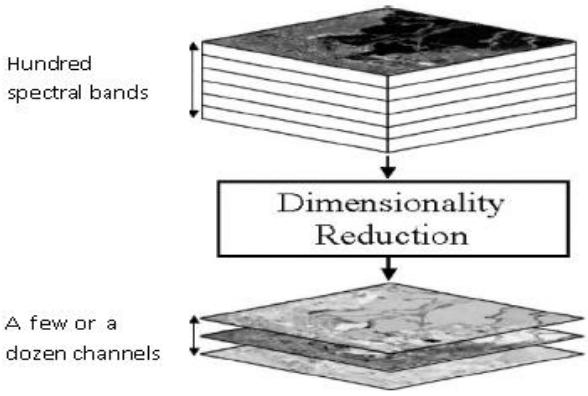

Fig 1: Preprocessed dimensionality reduction prior to information derivation from hyper spectral imagery.

Dimensionality reduction techniques are classified into linear and non-linear techniques. A linear method transforms the high dimensional data into a lower dimensional subspace using a linear mapping. Linear dimensionality reduction techniques finds a lower k dimensional representation S for the original p dimensional data X, where k < p and each variable of the reduced representation is a linear combination of the original variables as given in (1) or in (2).

$$S_i = w_{i,1}\, x_1 + \ldots + w_{i,p}\, x_p, \qquad \text{for } i=1,2,\ldots,k \qquad (1)$$

$$S = W_X \qquad (2)$$

Where W_{k*p} is the linear transformation weight matrix (Imola K. Fodor, 2002).

Among the previously investigated methods for linear dimensionality reduction, the most popular are: Principal Component Analysis (PCA) (Aleix M. MartoAnez et al, 2001) and Linear Discriminant Analysis (LDA) (Peter N. Belhumeur et al, 1997). The linear techniques are simple and easy to implement. In some cases, linear methods fail to find the intrinsic structure of non-linear datasets. To overcome this limitation, various non-linear methods like Locally Linear Embedding (LLE) (S.T. Roweis et al, 2000), Isomap (Joshua B. Tenenbaum et al. 2000) have been introduced.

The non-linear methods handle the complex non-linear structures efficiently. Sometime, the observed data lies on an abstract space called manifold of the original high dimensional space. For example, the variations that occur in some datasets can be represented using a low dimensional face manifold, lying on a high dimensional image space. This form of non-linear dimensionality reduction methods are known as manifold learning techniques (Joshua B. Tenenbaum et. Al, 2000). Manifold learning extracts the non-linear structures or low-dimensional manifolds embedded in the high dimensional data sets and this low dimensional embedding preserves the structures that exist in the original space. In such cases, the non-linear dimensionality reduction techniques can find the manifold structure efficiently, whereas the linear techniques fail to do so.

The remainder of this paper is organized as follows. Section 2 describes about face image retrieval system, section 3 presents related works and conclusion is described in section 4.

2. FACE IMAGE RETRIEVAL SYSTEM

In the recent years, automatic face recognition has become an active research topic in the field of pattern recognition. The recognized faces find their application in visual surveillance, retrieval of an identity from a face database for forensic applications, criminal investigations and security systems.

With regard to the machine learning approach, image retrieval can be viewed as the problem of classifying images into different classes. In this context, the performance metric is classification accuracy. The retrieval rate directly depends on this classification accuracy. Classification accuracy is a widely used metric for evaluation of machine learning systems. A face recognition system consists of two major steps.

- Feature Extraction
- Classification

The feature extraction methods extract the features for the optimal representation of the faces belonging to a class and then separate faces from different classes. Normally, the feature extraction is done by representing the data in a lower dimensional space computed using linear or non-linear dimensionality reduction methods. After the features are extracted, the classifier is trained for training images and finally, classifier classifies the face images to the appropriate classes.

3. RELATED WORK

Recently a significant number of works have been done on face detection and recognition. Face recognition methods can be classified as feature based methods and subspace methods. Feature-based approaches extract the local features such as the position of the eyes, nose, mouth etc. Subspace method reduces the dimension of the data, while retaining the maximum separation between distinct classes.

'Eigenface' (M.Truk et al, 1991) and 'Fisherface' (Peter N. Belhumeur et. Al, 1997) are the two widely used subspace methods for face recognition. The 'Eigenface' approach uses the linear unsupervised dimensionality reduction method Principal Component Analysis (PCA) for subspace generation, whereas the 'Fisherface' approach uses linear supervised dimensionality reduction method Linear Discriminant Analysis (LDA).

PCA finds a projection on a lower dimensional representation, where along the principal component most of the data variation occurs in an unsupervised manner. PCA provides the accurate representation of the data with minimum reconstruction error and also finds the best axis for projection (Peter N. Belhumeur et. Al, 1997). The main aim of LDA is to maximize the discrimination between different classes, while minimizing the within class distance. In classification systems, LDA is superior to PCA because, it provides higher class discrimination by using the class information (Aleix M. MartoAnez et. Al, 2001) and so, LDA is widely used in face recognition systems. Also, the experiment results show that when the number of samples per class is small, PCA might outer perform LDA.

LLE and Isomap are the non linear dimensionality reduction techniques that has been used for face recognition (S.T. Roweis et al. 2000) (A.Zagouras, 2007). LLE is suitable for processing the high dimensional data that contains non linear structures and the key merit of LLE is that it preserves the neighborhood structure of the high dimensional data. The discrimination capability of LLE is less compared to LDA (A.Zagouras, 2007). Inorder to recover the true dimensionality

and geometric structure of the high dimensional data, Isomap is widely used (Joshua B. Tenenbaum et. Al, 2000). Isomap retains the global structure but, computationally expensive for large datasets. Experiments show that, when Isomap is applied for face recognition, it has the worst performance (A.Zagouras, 2007). Furthermore, many work has been done in the literature by combining different dimensionality techniques for face recognition (Junping Zhang et al, 2004). However, the classification accuracy of such combinations is not so high.

Some works (Mahesh Pal et al., 2010) explored about the need of dimensionality reduction as a preprocessing step to classification. Support Vector Machine (SVM) is a widely used method for classification. SVM constructs a hyper plane in which the margin between different classes should be high and so, it provides better classification accuracy. Experiments are conducted for classification of hyper spectral data using SVM. Experimental results show that the classification accuracy of SVM can be increased by reducing the dimensionality of the data (Mahesh Pal et al., 2010). Thus, the authors proved that, dimensionality reduction is an essential pre-processing stage for classification by SVM. By the addition of features, the classification accuracy of SVM decline significantly and this is best apparent for small training sets. However, even for large training data set, dimensionality reduction have an important role. Experimental results show that there is a strong dependence between dimensionality reduction and classification accuracy of SVM. When the feature dimensions considered were 55, 60 and 65, the classification accuracy was 92.24%, 92.11% and 91.76%. Besides providing classification accuracy, dimensionality reduction offers the following advantages to SVM:

- Improved speed of the classification by reducing feature set size
- Reduction in data storage requirements

.A method for face recognition system based on the combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for feature extraction is introduced in the literature (Jianke Li et al., 2009). In order to perform feature extraction, a commonly used method is PCA and it provides better dimensionality reduction. After performing the PCA transformation, the resulting components are uncorrelated and have less reconstruction error. LDA increases the class separation and extracts the LDA features, which is advantageous for classification. So, for providing higher discrimination of the samples, LDA is used after PCA transformation. Thus PCA forms a PCA subspace and the combination of PCA and LDA form the LDA subspace having high discrimination. For classification purpose, Nearest Neighbor Classifier (NNC) is used. Experiments are conducted on the ORL face database containing 400 images of 40 individuals. Classification accuracy for different training sample size and for various feature dimensions was calculated. An increase in training sample size causes an increase in the classification accuracy of both PCA method and combined PCA and LDA method. Also, it can be concluded that the PCA and LDA combination method provides better classification accuracy than the PCA alone method for face recognition (Jianke Li et al., 2009).

Another work, (Md. Omar Faruqe et al, 2009) proposed a face recognition system using Principal Component Analysis (PCA) and support vector machine (SVM). Since the major components of a face recognition system are feature extraction and classification, its performance strongly depends on the factors including the way in which the features are extracted, how accurately the features are classified into a group and the selection of a classifier. Therefore, the selection of the feature extractor and the classifier is very important in a face recognition system. In this work, the authors have used PCA for feature extraction and SVM for classification.

In SVM based classification systems, the classification is done by the construction of a new hyper plane that provides maximum separation between the data item that is, the margin of the hyper plane should be maximum. The vectors which are lying near to the hyper plane are called support vectors. Every point in the input space is non-linearly mapped into a high dimensional feature space. This mapping is done using kernel functions. Experiments are conducted on the ORL face database containing 400 images of 40 individuals. Randomly selected 200 samples serves as the training set. This sample set is used for constructing Eigen faces and also for training the SVM. A change in feature vector size causes a change in input vector size of the classifier because; the output after PCA transformation is given directly to the classifier. From the experiments conducted it can be concluded that, PCA and SVM combination provides better classification accuracy for face recognition on the ORL face database.

A new dimensionality reduction technique using support vector machine is introduced in the literature (Sangwoo Moon et al, 2010). The decision vectors which are used for classification purpose in SVM are used as the mapping vectors for dimensionality reduction. By using these mapping vectors, it produces a feature set having high efficiency and better classification ability. Normally, the mapping vectors which are derived from the SVM is highly redundant. So, further reduction of redundancy is done based on a factor called asymmetric decorrelation measure based on angular distance between two mappings and the quality of mapping, which is obtained from its classification performance. Redundancy removal process eliminates the less meaningful mapping vectors. The mapping vectors which are survived after the removal process are far away from each other and also provide higher classification accuracy. Experiments are conducted on the handwritten numeric characters dataset. The feature dimensionality of SVM was 30 and the classification accuracy was 98.2%. It can be concluded that, SVM provides better dimensionality reduction as well as higher classification accuracy.

A summary of the different dimensionality reduction methods used for face recognition in the literature survey is given in Table 1. Experiments are conducted on the ORL database containing 400 images of 40 individuals. The images were taken at different times, varying lighting slightly, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the images are taken against a dark homogeneous background.

Table 1. Classification accuracy Vs Feature dimension

Method	Classifier	Feature Dimension	Classification Accuracy
PCA	NNC	40	85%
PCA + LDA	NNC	40	93%
PCA	SVM	40	97%

From the literature survey, it is observed that whatever be the classifier used for classification, dimensionality reduction is an essential step prior to the classification stage (Mahesh Pal et al, 2010). A classification system provides higher classification accuracy only if it is preceded by a better dimensionality reduction stage and better class discrimination stage. When the Nearest Neighbor Classifier (NNC) is used for classification of face images in ORL database, PCA alone method provided a classification accuracy rate of 85%, which is satisfactory. But, this rate is significantly improved into 93% by using PCA and LDA combination. It is clear that, combination of PCA and LDA is very much better than PCA alone for improving classification accuracy of SVM (Jianke Li et al., 2009).

4. CONCLUSION

Studies show that, the feature dimensionality has an adverse impact on the classification accuracy of a classifier. SVM is a widely used classifier. Therefore, it is beneficial to perform a dimensionality reduction method prior to the SVM. PCA is found to be a suitable dimensionality reduction method for SVM. By providing better dimensionality reduction and higher class discrimination prior to the classification process leads to higher classification accuracy. PCA is an extensively used dimensionality reduction method. But, the class discrimination capability is limited. It demands a dimensionality reduction method with high class discrimination. LDA gives higher class discrimination among the different classes. By the induction of LDA between PCA and SVM improves the classification accuracy. Combination of PCA and LDA is used for improving the capability of LDA when a few samples of images are available. It proposes the combination of PCA and LDA prior to SVM improves classification accuracy. During the dimension reduction using PCA, main features that are important for representing face images are extracted , LDA selects the significant features for class separability and SVM classifier that classify input face images into one of available classes. Experiments are conducted on the Olivetti Research Laboratory (ORL) face database. The method improved the classification accuracy of the system into a rate of 98.35% for a feature dimension of 20 as shown in Table 2. Thus, a face retrieval system using the combination of PCA and LDA for feature extraction and SVM for classification is a suitable method that provides better results compared to the other methods in the literature.

Table 2. Classification accuracy Vs Feature dimension for the proposed method

Feature Dimension	Classification Accuracy
11	96.6
20	98.3
30	97.5
40	97.5

5. REFERENCES

[1] Sergio A. Alvarez, “An exact analytical relation among recall, precision, and classification accuracy in information retrieval", 2002

[2] G. Chen and S.E. Qian, “ Dimensionality reduction of hyper spectral imagery using improved locally linear embedding ” , Journal of Applied Remote Sensing 1, 2007.

[3] Imola K. Fodor, “A Survey of Dimension reduction techniques”, June 2002.

[4] Aleix M. MartoAnez and Avinash C. Kak, “PCA versus LDA”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, February 2001.

[5] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, NO. 7, July 1997

[6] S.T. Roweis and L. Saul, “Nonlinear dimensionality reduction by locally linear embedding", Science, vol 290 22, No. 5500, December 2000, pp. 2323-2326.

[7] Joshua B. Tenenbaum, Vin de Silva and John C. Langford, “A Global Geometric Framework for Nonlinear Dimensionality Reduction", science, VOL 290 22, December 2000.

[8] Mahesh Pal and Giles M. Foody, “Feature Selection for Classification of Hyperspectral Data by SVM", IEEE Transactions on Geoscience and Remote Sensing, Vol. 48, No. 5, May 2010

[9] Jianke Li, Baojun Zhao and Hui Zhang, “Face Recognition Based on PCA and LDA Combination Feature Extraction", The 1st International Conference on Information Science and Engineering, 2009

[10] Md. Omar Faruqe and Md. Al Mehedi Hasan, “Face Recognition Using PCA and SVM", The 3rd International Conference on Antiounterfeiting, Security, and Identi_cation in Communication, 2009

[11] Sangwoo Moon and Hairong Qi, “Effective Dimensionality Reduction based on Support Vector Machine", International Conference on Pattern Recognition, 2010

[12] A.Zagouras, A.Macedonas, G.Economou and S.Fotopoulos, “An application study of manifold learning-ranking techniques in face recognition", Multimedia Signal Processing, 2007, Pages: 445 – 448

[13] Junping Zhang, Huanxing Shen and Zhi-Hua Zhou, “Unified Locally Linear Embedding and Linear Discriminant Analysis Algorithm for Face Recognition", Sinobiometrics, LNCS 3338, pp. 269-304, Springer 2004

[14] M.Truk and A. Pentland,” Eigen faces for recognition”, Journal of CognitiveNeuroscience,3:72-86,1991